ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/385753906

Environmental DNA barcoding reveals general biodiversity patterns in the large tropical rift Lake Albert

Article *in* The Science of The Total Environment · November 2024

DOI: 10.1016/j.scitotenv.2024.177308

CITATIONS
0

READS
11

13 authors, including:

Julius Tumusiime
Mbarara University of Science & Technology (MUST)
17 PUBLICATIONS **94** CITATIONS
SEE PROFILE

Francis Ssenkuba
Mbarara University of Science & Technology (MUST)
3 PUBLICATIONS **1** CITATION
SEE PROFILE

All content following this page was uploaded by Julius Tumusiime on 13 November 2024.

The user has requested enhancement of the downloaded file.

Science of the Total Environment 957 (2024) 177308

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Environmental DNA barcoding reveals general biodiversity patterns in the large tropical rift Lake Albert

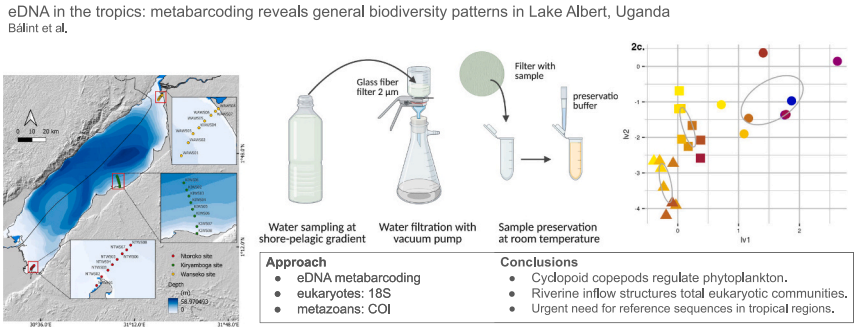
Miklós Bálint [a,b,c,*], Julius Tumusiime [d,e], Justine Nakintu [d], Damian Baranski [c], Leonie Schardt [c], Juliane Romahn [a,b,c], Marie-Claire Dusabe [e], Casim Umba Tolo [d], Grace Rugunda Kagoro [d], Francis Ssenkuba [d], Annett Junginger [f], Christian Albrecht [d,e]

[a] Institute of Insect Biotechnology, Justus-Liebig University, Heinrich-Buff-Ring 26, 35392 Gießen, Germany
[b] Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[c] LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[d] Mbarara University of Science and Technology, P.O Box 1410, Mbarara, Uganda
[e] Institute of Animal Ecology and Systematics, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Gießen, Germany
[f] Department of Geosciences, Eberhard Karls University, Hölderlinstr. 12, 72074 Tübingen, Germany

HIGHLIGHTS

- eDNA is well preserved in large, warm tropical freshwaters.
- eDNA metabarcoding confirms patterns in tropical plankton distribution.
- Comprehensive reference sequences are urgently needed.
- eDNA already allows efficient taxonomy-free biomonitoring in the tropics.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Beatrice Opeolu

Keywords:
Tropical freshwater ecosystems
Reference sequences
Plankton
Biodiversity gradients
eDNA
Metabarcoding

ABSTRACT

Lake Albert, Africa's seventh-largest lake and a biodiversity hotspot, faces significant environmental challenges, including unregulated anthropogenic pressure and a lack of comprehensive biological studies. To address the scarcity of biodiversity data, we utilized environmental DNA (eDNA) metabarcoding to assess the lake's eukaryotic and metazoan communities. Surface water samples were collected at three distinct locations: close to the southern inflow of the Semliki River, the central part of the lake, and close to the northern inflow of the Victoria Nile and outflow of the Albert Nile. We aimed to study ecological patterns across the lake, focusing on sequence variant richness and community composition, testing for differences among locations and between shoreline and pelagic zones. Consistent with previous morphology-based observations, our results revealed differences in community composition among the three sites, with cyclopoid copepods dominating the communities. Distance from shore was a significant factor influencing community composition, confirming expectations about the effects of nutrient and oxygen availability gradients. However, the lack of comprehensive reference sequence data limited accurate taxonomic assignments. Despite these limitations, our study

* Corresponding author at: Institute of Insect Biotechnology, Justus-Liebig University, Heinrich-Buff-Ring 26, 35392 Gießen, Germany.
E-mail address: miklos.balint@senckenberg.de (M. Bálint).

https://doi.org/10.1016/j.scitotenv.2024.177308
Received 9 August 2024; Received in revised form 13 October 2024; Accepted 28 October 2024
Available online 13 November 2024
0048-9697/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

demonstrates that eDNA metabarcoding is highly useful for assessing biodiversity in underexplored tropical freshwater ecosystems. We advocate for urgent efforts to generate reference sequences from tropical regions to enhance the utility of eDNA for biodiversity monitoring and conservation. Our findings underscore the potential of eDNA in providing insights into ecological patterns of entire communities and emphasize the need for comprehensive studies addressing the full taxonomic spectrum in tropical freshwater ecosystems.

1. Introduction

Africa's freshwater systems are widely acknowledged as biodiversity hotspots, yet they face significant challenges. Despite the high levels of biodiversity present, there is increasing unregulated anthropogenic pressure on these ecosystems, leading to reported declines in species populations and even the likelihood of many species facing a high threat of extinction (IPBES, 2019). Another major challenge is the taxonomic crisis, with many taxa in African freshwaters remaining unstudied or unknown. In order to address these issues, one approach that has gained traction in studying entire aquatic organism communities is environmental DNA (eDNA) analysis (Schenekar, 2023). While this method has been somewhat established in extratropical regions for monitoring and assessing biodiversity, Africa significantly lags behind in adopting this approach (Takahashi et al., 2023).

Tropical freshwaters are dramatically underrepresented in eDNA studies (von der Heyden, 2023). This is particularly true for Africa, with only six eDNA studies reported until 2023 from all of the continent's hyperdiverse freshwater ecosystems (Schenekar, 2023). The focus of the handful of existing studies is either on disease vectors and human pathogens like parasitic flatworms of the genus *Schistosoma* (Sengupta et al., 2022) in Lake Tanganyika (Doble et al., 2020) and River Okavango (von der Heyden et al., 2023). There is a notable lack of comprehensive studies covering the entire taxonomic spectrum in large lakes or rivers (Takahashi et al., 2023), including all of the African Great Lakes, well-known centers of tropical freshwater biodiversity (Odada and Olago, 2006; Salzburger et al., 2014). At the same time, human action, including anthropogenic climate change is severely threatening these delicate systems (Salzburger et al., 2014). Considering that rift lakes are characterized by high levels of endemism, researching these systems is crucial. Given these research gaps, there is a pressing need for comprehensive eDNA studies which address the full taxonomic spectrum of African freshwater ecosystems. Further, establishing present-day biodiversity baselines using eDNA in these areas can help to track changes over time, offering critical insights for conservation and sustainable management, given that historical biodiversity data is largely lacking from these regions.

One such ecosystem that exemplifies the need for detailed study is Lake Albert. The lake is Africa's seventh largest lake (Ali and Abd Ellah, 2023), located in the western arm of the African Rift System at 619 m asl (Fig. 1). It is sitting on major sediment deposits of Miocene origin, but it is relatively young in geological terms (i.e. Pleistocene). The lake experiences major lake level fluctuations, and it was likely even completely dried out during the Pleistocene (Berke et al., 2014; Beuning et al., 1997). It is a typical rift lake, ca. 150 km long and 35 km wide on average with a maximum depth of only 56 m and covering 5600 km^2

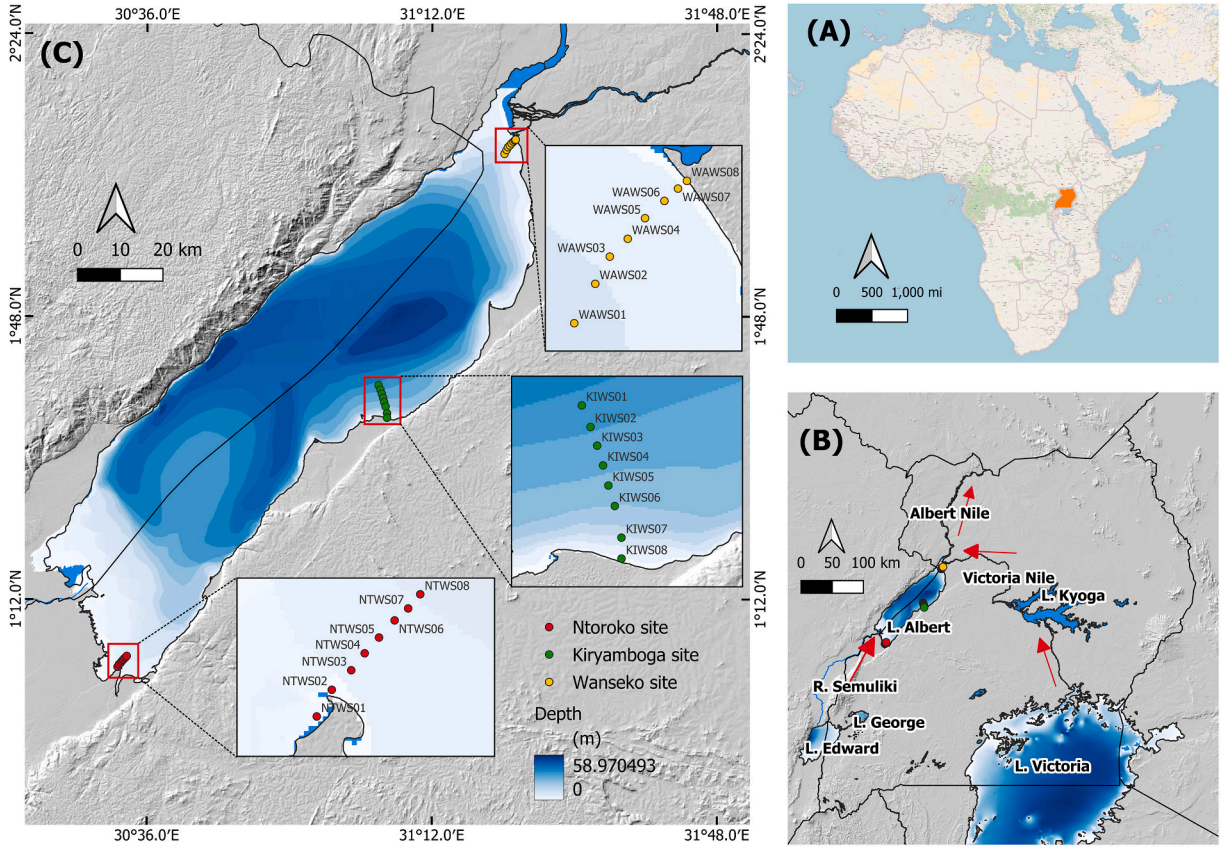


Fig. 1. A) Location of Uganda in Africa, B) sampling locations in Uganda, C) detailed sampling points on Lake Albert. Samples were collected from the southern (Ntoroko), central (Kiryamboga) and northern (Wanseko) parts of the lake along distance gradients from the shore. Red arrows mark the flow directions of the Semliki River, Victoria Nile and Albert Nile.

(Green, 2009). Mixing of its water column is pronounced during several months (Green, 2009). This is reflected in its mesotrophic state (Morana et al., 2023) and relatively high oxygenation down to the profundal (Beadle, 1981). Lake Albert's water level is mostly balanced by the Nile River which flows in and out of the lake at its northern tip (Beadle, 1981). Another major tributary is the Semliki River in the south, connecting from the Lake Edward-George system. These two rivers largely determine the current limnological conditions in Lake Albert (Beadle, 1981; Nakiyende et al., 2023), although there are some minor tributaries from the Rwenzori Mountains to the west of the lake. The lake is mesotrophic, dominated by non-humic primary production (Morana et al., 2023).

Lake Albert supports an immense range of aquatic biodiversity, including macro- and micro-invertebrates, fish, zooplankton, reptiles, birds, aquatic macrophytes, and mammals (Wandera and Balirwa, 2010). The lake is notable for endemisms, particularly with fish species such as *Lates macropthalmus* and *Engraulicypris bredoi (*Froese and Pauly, 2024*)*, and molluscs like *Bellamya rubicunda, Biomphalaria stanleyi*, and several species of *Gabbiella* (Brown, 1994). This highlights the urgent need for a comprehensive biodiversity inventory of the lake to inform sustainable management strategies and protect its diverse ecosystems. However, knowledge of the lake's fauna and flora is still scant, as information collection started around the middle of last century (Verbeke, 1957). As part of the Upper Nile freshwater ecoregion, nilotic elements dominate the fauna and flora in these reports. The few biodiversity studies concentrated on fish as Lake Albert supports the most diverse commercial fisheries in Uganda with at least 55 species (Wandera and Balirwa, 2010), although with comparatively low degree of fish species endemism (Nakiyende et al., 2023). Another prominent group of organisms that has been studied in this lake are molluscs. Some gastropod groups have received more attention due to their role as intermediate hosts for schistosomiasis (Rowel et al., 2015; Tumusiime et al., 2024). Knowledge about the lake's phytoplankton comes from a one-year sampling campaign (1961–1962), published in 1997 by Evans. The data shows that nutrient conditions in the lake are largely influenced by hydrological inflows from the two main rivers. The influence of the Victoria Nile in the north is limited, as it flows out of the lake almost immediately after its inflow. Nonetheless, the Nile has a considerable impact on the lake at its northernmost end (Beauchamp, 1956). The Semliki River in the south is a major source of nutrients, especially in the southern parts of the lake. This results in an end to end compositional polarization of the phytoplankton, with the riverine inflows of Semliki and Nile creating more eutrophic conditions, compared to the more oligotrophic free pelagic zone in the mid-lake center (Evans, 1997). Interestingly, phytoplankton biomass decreased between the 1960s and the 1990s, likely as a result of changes in wind conditions (Lehman et al., 1998). There is little information available about the zooplankton and macro-invertebrate communities which are key contributors to the food web and ecosystem functioning of aquatic systems. Green (1967) focused on fish predation of a *Daphnia* species and reported a reduction of fish predation on zooplankton in the mid sections of the lake. Lehman et al. (1998) report an overall dominance of cyclopoid copepods in Lake Albert's zooplankton, with a much smaller contribution from calanoids and cladocerans. Rotifers were found to be rare in the lake (Lehman et al., 1998), in comparison with other African lakes (Green, 2009).

Gradients from shoreline to pelagic conditions have been detected in Lake Albert, where shoreline waters experience more temporal oscillations, likely affecting water movement and stratification (Talling, 1963). This variability is reflected in distinct phytoplankton biomasses between inshore and offshore areas (Lehman et al., 1998). Similarly, the composition of zooplankton differs between pelagic and shoreline zones in Lake Albert (Green, 1971). (Green, 1967) even observed a morphological shift of *Daphnia lumholtzi* along the shore-pelagic gradient, attributed to feeding pressure from planktivorous fish. Compositional differences of plankton between shoreline and pelagic areas likely result from variations in nutrient concentrations and other biological or environmental factors (Chomicki et al., 2022; Dusabe et al., 2024; Green, 1967).

However, records of all organism groups are sporadic at best (and inexistent at worst) in Lake Albert, with many organism groups completely unstudied. This highlights a significant gap in understanding its ecosystem dynamics. Consequently, we likely face a massive underestimation of the actual degree of biodiversity and endemism. At the same time, Lake Albert faces serious biodiversity threats with anticipated extinctions, especially now that the lake ecosystem is exposed to oil exploration risks (Verheyen et al., 2016). The oil exploration activities came with infrastructural developments i.e. improved road networks which have accelerated rural-urban migration into this region exacerbating the problem (Ericson, 2014). A biodiversity crisis has been identified as the lake's most important recent challenge (Nakiyende et al., 2023), threatening its fisheries. Other important stressors include a destruction of shoreline vegetation, and increasing water pollution (Nakiyende et al., 2023). Considering that the lake is part of the Nile basin, its anthropogenic degradation can impact ecosystem health far downstream. However, the methodological and general knowledge gaps hamper conservation efforts and sustainable management ((Nakiyende et al., 2023)).

Environmental DNA is a highly promising approach for large scale biodiversity surveys in neglected tropical freshwaters, and to complement the so far sparse records of biodiversity. Environmental DNA might also become an important tool for monitoring the lake's biodiversity, contributing to a sustainable use and management of its ecological resources. However, the use of eDNA in large tropical lakes might encounter several issues. Conditions in the water column of tropical lakes are very distinct compared to temperate lakes, with high temperatures throughout the water column resulting in frequent mixing (Talling, 1969). This mixing causes nutrient redistribution and deep oxygen penetration (De Crop and Verschuren, 2021). Hydrological differences might translate into potential preservation issues for eDNA. DNA might decompose faster at high temperatures (Lamb et al., 2022), and available oxygen supports microbial activity, which might result in rapid DNA removal from the water column (Bairoliya et al., 2022). Finally, even if DNA is preserved and suitable for community assessments, identification problems arise due to the lack of comprehensive sequence databases. Very few sequence data exist for tropical regions (Lopes-Lima et al., 2024), posing a key problem for African eDNA studies (von der Heyden, 2023). An explicit evaluation of this problem is still lacking. Altogether, the suitability of eDNA must be evaluated for tropical lakes.

Our specific objectives were to evaluate the suitability of eDNA for describing general biodiversity patterns in tropical large lakes. Specifically, our study aimed to achieve three primary objectives. First, we sought to test environmental DNA (eDNA) methodologies by examining both previously observed and expected patterns within plankton communities of Lake Albert. Second, we aimed to evaluate the feasibility of making accurate taxonomic assignments using existing sequence references, thereby assessing the reliability of current genetic databases for identifying tropical freshwater biodiversity. Third, we wanted to generate the first comprehensive, community-wide eDNA baseline dataset of plankton diversity from any of the African Rift Lakes. This baseline is intended to serve as a crucial reference point for future biodiversity monitoring and conservation efforts in these unique and ecologically significant regions.

Along these aims we hypothesize that 1) eDNA reflects previous findings on plankton composition, 2) eDNA is sensitive enough to detect expected shoreline - pelagic gradients, and 3) currently available sequence data is insufficient for a reasonable taxonomic assignment of African lake plankton biodiversity.

2. Materials and methods

2.1. Sampling

Permission to collect and transfer samples was granted by Uganda National Council of Science and Technology (UNCST) under permission number NS148ES. Sampling was conducted in March 2023 at three distinct locations within the lake: the southern region near the inflow of the Semliki River (Ntoroko District), which carries water from Lake Edward; the central part of the lake (Kiryamboga) in the Tonya area (Hoima District); and the northern area, Wanseko (Buliisa District), close to the inflow and outflow of the White Nile (Victoria Nile and Albert Nile, respectively). Eight surface water samples were collected at each location along a transect from a boat, spaced approximately 500 m apart, starting from the deepest end of the transect (Fig. 1, Supp. Table 1). Samples were collected into plastic bottles and filtered using glass fiber filter membranes (2 μm pore size, thickness 380 μm, Ø 47 mm, Sigma-Aldrich, Darmstadt, Germany) with a reusable bottle-top filter, fitting 45 mm bottle threads (Fisher Scientific, Schwerte, Germany). The filter was operated manually with a vacuum pump. After filtration, the glass fiber membranes were folded and stored in 2 ml test tubes, preserved with Zymo DNA/RNA Shield buffer (Zymo Research, Freiburg, Germany). These tubes were kept at ambient temperature in the field (up to two weeks) and subsequently frozen at −20 °C in the laboratory prior to DNA extraction.

2.2. DNA extraction

2.2.1. Laboratory preparations before DNA/RNA extractions

PCR workstations (VWR International GmbH, Darmstadt, Germany) were cleaned with a 10 % dilution of a bleach-based cleaner (DanKlorix, CP GABA GmbH, Hamburg, Germany), followed by distilled water before every use. Pipettes in the workstations were cleaned with DNA/RNA-ExitusPlus IF (ITW Reagents, S.R.L., Monza, Italy) before every use. The designated RNA workstation and pipettes were additionally cleaned with RNAse AWAY (Molecular BioProducts, Thermo Fisher Scientific, Massachusetts, USA) before every use. These cleaning steps were repeated once the lab work in the boxes was done for the day. Prior to the lab work starting and after finishing, both boxes had the internal UV lights running for 30 min. Forceps were coated in 70 % denatured ethanol and then sterilized with fire. Pipette tip boxes were put into a shortwave UV crosslinker with open lids for 10 min each.

2.2.2. Filter preparations and lysis

Samples were randomly assigned to four different extraction batches (Supp. Table 1) to avoid batch effects (Bálint et al., 2018). These batches were processed on four separate days in June 2023. Filters were removed from storage solution tubes using sterilized forceps (changed between every sample tube) and unfolded on a stainless steel surface in the UV workstation. One-third of the filter membranes was cut into small pieces with scissors and placed in a bashing bead lysis tube of the ZymoBIOMICS DNA/RNA Miniprep kit (Cat. nr. R2002, Zymo Research Europe GmbH, Freiburg, Germany). Additionally, 400 μl of the storage buffer from the same sample was transferred to the lysis tube. The unused portion of the filter membranes was placed back in the original sample tube (still containing leftover Shield buffer) and kept at −20 °C as a backup. A negative extraction control was run alongside each extraction batch (two controls for the last extraction batch), consisting of 400 μl of DNA/RNA Shield buffer transferred to a lysis tube. Lysis tubes were shaken for 40 min on a Vortex Genie 2 using a horizontal tube adapter, set to maximum speed (3200 rpm). The stainless steel surface was cleaned between processing every sample with 10 % bleach solution, and rinsed with distilled water. Scissors were submerged into a 10 % bleach solution for a few minutes and rinsed with distilled water prior to processing the next filter.

2.2.3. DNA/RNA extraction

DNA and RNA were extracted in parallel from the cut and lysed filter membranes using the ZymoBIOMICS DNA/RNA Miniprep kit (Cat. nr. R2002, Zymo Research Europe GmbH, Freiburg, Germany), following the manufacturer's instructions. DNA was eluted in 80 μl of AccuGENE molecular biology grade water (Lonza, Verviers, Belgium); RNA was eluted in 50 μl of molecular biology grade water. DNA extracts were subsequently stored at −20 °C until further processing; RNA extracts were subsequently stored at −80 °C.

2.3. PCR reactions and sequencing

Metabarcoding PCR reactions were automated using a Biomek i7 workstation (Beckman Coulter GmbH, Germany) in a molecular prePCR laboratory. To avoid contamination, several measures were taken to sterilize the laboratory equipment and all surfaces before and after pipetting. All surfaces were treated with DNA Exitus (Kisker, Steinfurt, Germany), bleach and UV light, and plastic equipment was sterilized under UV light. In addition, protective equipment (full body suits, face masks, shoe covers, face masks, etc.) was used. DNA extracts were amplified with a) the Euka02 primers (F: 5'-TTTGTCTGSTTAATTSCG-3′, 5'-CACAGACCTGTTATTGC-3′, (Guardiola et al., 2015)) and b) the miCOIint/jgHCO2198 primers (F: 5′-GGWACWGGWTGAACWGTWTAYCCYCC, R: 5′-TAAACTTCAGGGTGACCAAARAAYCA), (Geller et al., 2013; Leray et al., 2013) using AmpliTaq Gold™ Mastermix (Thermo Fisher, Waltham MA, USA). Individually labeled PCR replicates of the Lake Albert samples were pooled with reactions from three other unrelated projects. We followed a one-step-PCR approach in which every sample had a unique nucleotide index combination of both primer pairs of 11 base pairs. This allowed three base pair mismatches for index identification (Taberlet et al., 2018). PCR was performed using 10 μl AmpliTaq GOLD Reaction buffer, 2 μl GC Enhancer, 3 μl water, 2 primer pair mixture (5 μM), 3 μl sample. The PCR positive control for Euka02 was DNA extract of a single Nematoda species (*Rhomborhabditis regina*, 800 individuals); the PCR positive control for the metazoan primer pair was a mock community containing DNA extracts of 24 species. The metabarcoding PCRs were run with eight PCR negative controls, 24 multiplexing negative controls and four PCR positive controls. The multiplexing negative controls were no-template no-mastermix control samples used to control for index jumps during the library preparation process (Taberlet et al., 2018). PCR reactions were performed in four individually indexed replicates for all samples and controls. The cycling program for Euka02 included an initial denaturation step at 95 °C for 10 min, 35 cycles of 95 °C for 30 s, 45 °C for 30 s, 72 °C for 60 s, and a final elongation step at 72 °C for 7 min. The cycling program for miCOIint/jgHCO2198 included an initial denaturation step at 95 °C for 10 min, 35 cycles of 95 °C for 30 s, 57 °C for 90 s, 72 °C for 60 s, and a final elongation step at 72 °C for 7 min. After amplification, PCR products were pooled and purified using the MinElute PCR Purification Kit (Qiagen). Library preparation and paired-end sequencing (NovaSeq 6000, 2 × 150 bp for Euka02 and Miseq 2 × 250 bp) was performed with the MetaFast protocol at Fasteris SA (Geneva, Switzerland).

2.4. Bioinformatics and taxonomic assignment

Bioinformatic analysis was conducted using the Obitools pipeline v4.0.5 (Boyer et al., 2016). The paired-end reads were assembled with a minimum overlap of ten bases and then demultiplexed into PCR replicates. These replicates were de-replicated into amplified sequence variants (ASVs). Sequences were discarded which had variants of other sequences with a count >5 % of their own count. We kept ASVs only if they had lengths between 70 and 350 bp, and if they were represented by at least 10 reads.

The ASV x PCR replicate matrix underwent further processing with a custom R script, following the recommendations of (Taberlet et al., 2018). To account for potential PCR or sequencing errors, ASVs with low

read abundance were removed if their total reads fell below a threshold (18S: 100 reads; COI: 75 reads) based on frequency criteria (Drake et al., 2022). Additionally, replicates with spurious richness were filtered out based on their richness - read profiles, specifically those with fewer than 7000 reads for 18S and fewer than 75 ASVs for COI. For quality control, if an ASV was present in any of the negative controls (including extraction, PCR, and multiplexing negatives), the maximum number of reads for that ASV in any negative control was calculated and subtracted from each replicate. Nonmetric multidimensional scaling (NMDS) ordination plots of the replicates were generated using the "vegan" package in R (v4.4.0) with Bray-Curtis dissimilarity. Replicates highly dissimilar to others from the same sample were removed. Subsequently, PCR replicates for each sample were summed into a community matrix, and altered observation depths due to replicate removal were corrected.

Taxonomic assignment of 18S ASVs was performed using local BLAST searches against the NCBI database, while COI ASVs were matched against all COI sequences from both NCBI and the Barcode of Life database (Meglécz, 2023). Only the best 100 BLAST hits with an e-value <0.001 were retained. The BLAST results were parsed with MEGAN v6 (Huson et al., 2016), and each ASV was assigned to its lowest common ancestor based on the top 10 % of bit scores of BLAST results.

2.5. Statistical analyses

All statistical analyses were performed in R v4.4.0 (R Core Team, 2024). Unequal water sample volumes and unequal sequencing depths of replicates were explicitly accounted for in all analyses by including them into models as predictors. This has three benefits in comparison to simple rarefactions: 1) avoids data loss, 2) allows remove their effects when using model-based analysis tools, and 3) explicitly quantifies the effect of biases linked to observation depth, in comparison to the effect of predictors of interest (reviewed in (Bálint et al., 2016)). We considered only the taxonomic groups specifically targeted by the primers in the analyses of community properties. This means that all eukaryotic ASVs were considered in the analyses which were amplified with the primers euka02, but all other ASVs (e.g. a few ASVs assigned to prokaryotes, or not assigned to anything known) were omitted. We considered only metazoan ASVs amplified with the COI primers, but all other ASVs, e.g. algae, were omitted.

2.5.1. ASV richness

With respect to ASV richness and community composition, we tested 1) the impact of inflowing rivers (three locations: Ntoroko - Semliki River, Kyriamboga - no major effluent, Wanseko - Victoria Nile, Fig. 1), and 2) the effects of a shoreline - pelagic gradient. ASV richness was modeled with four predictors (water sample volume, sequencing depth, sampling locality and distance from shore) in linear regressions.

2.5.2. Community composition

To select abundant and frequent ASVs, we modeled the log-normalized average read abundance of ASVs against their incidence with generalized additive models (mgcv package, (Wood, 2017)). We used GAMs primarily for exploring patterns, rather than for hypothesis testing (Simpson, 2018). GAMs are particularly suitable for this purpose because they do not impose a strict parametric relationship between variables, allowing for flexible pattern discovery. By doing so, we selected taxa with a high signal-to-noise ratio, enabling us to focus on taxa with meaningful biological variation rather than noise (Bálint et al., 2016). For analyses of 18S community composition, we retained abundant and frequent ASVs: these were represented by more than the average number of reads, and were present in at least six samples. For analysis of COI community composition, we retained frequent OTUs: these were present in at least eight samples. This resulted in sample x ASV matrices with 278 ASVs (18S), and 241 ASVs (COI).

For testing hypotheses about community composition, we employed generalized linear models (GLMs), which are well-suited for handling non-normally distributed data. Specifically, we modeled taxon read abundances using a negative binomial distribution to account for overdispersion in the data, as recommended for ecological counts (Lindén and Mäntyniemi, 2011). Although distance-based methods like PERMANOVA or NMDS are commonly used for community composition data, they rely on relative abundances, which may be biased due to technical factors like primer bias (Krehenwinkel et al., 2017). As such, we opted for GLMs within a joint species distribution model framework (Ovaskainen et al., 2017; Warton et al., 2015) to explicitly model the read counts per taxon, considering the underlying data generation process. The individual GLM results are summarized into community-level statistics at the end, providing an overview of driver importance of community-level differences in a format similar to results of distance-based approaches such as PERMANOVA (Wang et al., 2012). We consider this approach is better suited for metabarcoding data and allows for community-wide statistics to be calculated more robustly (Bálint et al., 2016).

We predicted community composition of frequent and/or highly incident ASVs with a model-based analysis of multivariate abundance data as implemented in the mvabund package (Wang et al., 2012). mvabund models the abundances of single taxa in a community against a set of predictors in a generalized linear modeling (GLM) framework, and summarizes model results into a community-level statistics. As ecological count data including sequencing reads are generally overdispersed, model-based community analyses are particularly helpful as they allow to include parameters about data properties (Warton et al., 2015). We fitted community models by specifying the distribution family as negative binomial using a log link, with an unknown overdispersion parameter. We created an analysis of deviance (ANOVA) table for the GLM, using a likelihood ratio test with Monte Carlo resampling (1000 bootstrap iterations).

We visualized compositional differences among the ASV communities with Bayesian ordination and regression as implemented in the boral package (Hui, 2016), by fitting the models with a negative binomial distribution. We parameterized the models with sample volume and sequencing depth as covariates, and accounted for variation unexplained by these covariates on two latent variables. Similarities and differences in ASV composition among samples were visualized along these latent variables in two dimensions, similar to a conventional ordination (Warton et al., 2015). We ran the MCMC sample for 40,000 iterations, discarded 10,000 iterations as burn-in with a thinning rate of 30.

3. Results

After the Obitools processing and subsequent R cleanup, the Lake Albert dataset contained 3,841,955 reads assigned to 4527 ASVs. The COI dataset contained 2,756,941 reads assigned to 1544 ASVs. Of the 4527 18S ASVs, 707 were assigned to algae, 330 to fungi, 839 to metazoans, 997 to non-algae protists, 36 to oomycetes, 65 to streptophytes, and 1553 could not be assigned to any taxonomic levels below the superkingdom of eukaryotes (Fig. 2, Supp. Table 2). Of the 1524 COI ASVs, 473 were assigned to algae, 16 to fungi, 87 to non-algae protists, 439 to other eukaryotes, and 509 to metazoans (Fig. 2, Supp. Table 2). Of the 4527 18S ASVs, 2874 (63.5 %) were shared across at least two locations, and 1653 (36.5 %) were observed at only a single location. Of the 509 metazoan COI ASVs, 321 (63 %) were shared across at least two locations, and 188 (37 %) were observed at only a single location.

3.1. Ecological patterns

3.1.1. ASV richness

The linear regression of 18S ASV richness had an adjusted $R^2 = 0.826$ ($p < 0.001$, $F = 22.89$, $df = 18$). The volume of water had no statistically significant effect. ASV richness statistically significantly increased with ~7 ASVs for every 10,000 reads ($p < 0.001$), and it was different

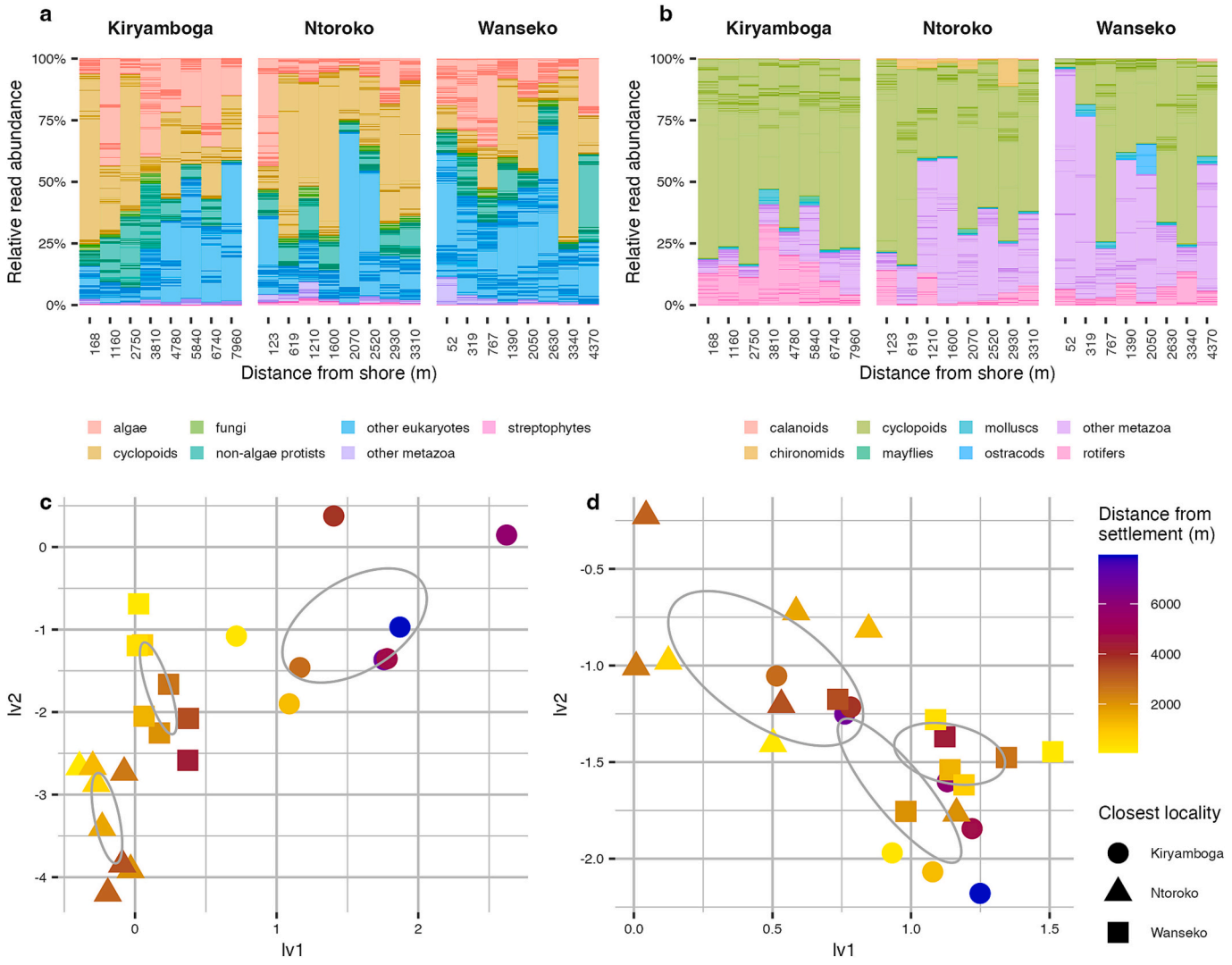


Fig. 2. Community composition recorded by eDNA. a) Dominant eukaryotic groups recorded by the 18S marker gene. b) Dominant metazoan groups recorded by COI. c–d) Model-based ordinations of eukaryotic (c) and metazoan (d) community composition.

between Kiryamboga and Wanseko ($p = 0.001$), but not between Kiryamboga and Ntoroko ($p = 0.169$). ASV richness statistically significantly decreased with ~6 ASVs for every 100 distance increase from the shore ($p = 0.002$).

The linear regression of COI ASV richness had an adjusted $R^2 = 0.705$ ($p < 0.001$, F = 12, df = 18). The volume of water had no statistically significant effect. ASV richness statistically significantly increased with ~5 ASVs for every 10,000 reads ($p < 0.001$), and it was different between Kiryamboga and Wanseko ($p = 0.007$), but not between Kiryamboga and Ntoroko ($p = 0.447$). ASV richness had no statistically significant relationship with the distance from the shore ($p = 0.445$).

3.1.2. Community composition

Composition of both 18S and COI ASV communities differed statistically significantly among locations (Table 1, Fig. 2). Of the 278 abundant and frequent 18S ASVs, 85 (31 %) had statistically significantly different read abundances among locations (Fig. 3, Supp. Fig. 1). These ASVs belonged to algae (20 ASVs), fungi (6) and protists (14). Many ASVs could not be identified beyond eukaryotes (45). Of the 241 frequent COI ASVs, 15 (6 %) had statistically significantly different read abundances among locations (Fig. 3, Supp. Fig. 2). These ASVs belonged to an annelid (likely an aquatic enchytraeid with 78 % sequence similarity to a *Limnodrilus* species), eight arthropods (two cyclopid ASVs, an amphipod, two insects, two branchipods and an ASV which could not be identified beyond arthropods), a chordate (likely a rodent, with ~79 % sequence similarity to a *Chinchilla* species), three snails and two rotifers (Supp. Fig. 2).

Table 1
ANOVA table of compositional differences of the model-based analysis of multivariate ASV abundances.

		Residual df	Df difference	Deviance	*p*
18S	(Intercept)	23			
	Sample volume	22	1	2101	0.001
	Sequencing depth	21	1	415	0.001
	Locality	19	2	3585	0.001
	Distance from shore	18	1	1701	0.001
COI	(Intercept)	23			
	Volume_filtered_ml	22	1	818	0.001
	seq_depth	21	1	360	0.001
	Closest_locality	19	2	1178	0.002
	dist_settlement_m	18	1	680	0.001

Composition of both 18S and COI ASV communities was statistically significantly affected by the distance from shore (Table 1, Fig. 2). Read abundances of statistically significantly affected 18S ASVs ($n = 42$) generally decreased with distance from the shore with the exception of a

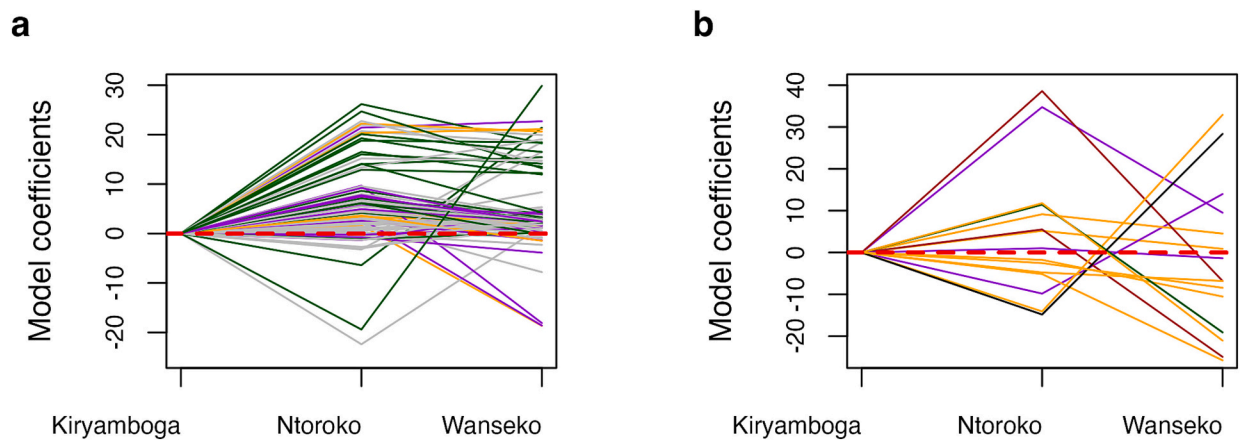


Fig. 3. Mean-centered model coefficients show the extent of differences in read abundance of ASVs at river-affected locations (Ntoroko: Semliki river, Wanseko: Victoria Nile), compared to a baseline (Kiryamboga, no major inflow). Most ASVs show higher read abundances at the river-affected locations. Black continuous lines represent ASV coefficients from the multivariate community GLM with statistically significantly different read abundances among the localities. Dashed red line marks the base of comparison. Colors mark major organism groups: a) 18S: green - algae; orange - fungi; purple - non-algae protists; gray - other eukaryotes. b) COI: green - annelids; orange - arthropods; chordates - black; molluscs - purple; rotifers - brown.

diatom ASV belonging to Coscinodiscophyceae (Supp. Fig. 3). The individual COI ASVs related to the distance from the shore belonged to arthropods, molluscs and rotifers. Read abundances of all statistically affected COI ASVs (n = 10) decreased with distance from the shore (Supp. Fig. 4).

3.2. Limitations of taxonomic identification

Average identity among 18S query and database sequences was 96 % (Fig. 4). Most assignments had low bit scores, even if sequence identity between query and database sequences was high (Supp. Fig. 5). Average identity among COI query and database sequences was 91 % (Fig. 4). The relationship between bit score and sequence identity was almost linear (Supp. Fig. 5).

4. Discussion

4.1. Environmental DNA confirms known and expected biodiversity patterns

4.1.1. Detected taxonomic composition and north-south differences

Most organisms recorded with the two marker fragments belonged to plankton (Fig. 2). This is not surprising given that we sampled only surface water. We found large read abundances of cyclopoid copepods in both 18S and COI data (Fig. 2). This confirms that the zooplankton communities are dominated by cyclopoid copepods, whereas calanoids and Cladocera represent <10 % of the total biomass (Lehman et al., 1998). Rotifers were also well represented as planktonic consumers in the eDNA data, especially at the central location of Kiryamboga (Fig. 2). Interestingly, previous accounts listed only a few rotifers, low compared to other lakes on the continent (Green, 2009). Rotifer abundance was reported to be highly localized with most species in the South of Lake Albert (Green, 2009). These differences may result from a reduction in phytoplankton biomass over recent decades, compared to the 1960s, likely due to diminished nutrient availability, which is itself a consequence of changes in wind mixing (Lehman et al., 1998). Additionally, reduced and/or selective predation pressure by planktivorous fish on zooplankton, attributable to the introduction of the Nile perch (Green, 1967; Lehman et al., 1998), might also play a role. Observed differences might originate from alterations in nutrient and temperature conditions since the 1990s, driven by increasing human impact and climate warming, as trophic conditions in the lake can shift within a matter of years or decades (Lehman et al., 1998).

Producers were mostly represented by diverse strains of algae, likely components of the phytoplankton, suggested by existing information on chlorophyll concentrations, which indicate algal photosynthesis to at least 25 m depth (Lehman et al., 1998). Interestingly, most of the ASVs with strongly different read numbers among the locations belonged to algae (Supp. Fig. 1). Read numbers of these algae were significantly higher at the river-influenced locations of Ntoroko and Wanseko, in comparison with the centrally located Kiryamboga. This is consistent with the only comprehensive data available (to our best knowledge) on the composition of phytoplankton communities from the lake, which dates back >60 years (Evans, 1997). This phytoplankton dataset reports an end to end polarization in the composition of phytoplankton, with higher concentrations at the two extremes (south and north) influenced by the Semliki and Nile rivers. Here the lake is more eutrophic, compared to the more oligotrophic free pelagic zone in the mid-lake center (Evans, 1997). Lake Albert is considered a rare case among tropical lakes as its planktonic communities are largely controlled by its hydrographic and hydrological character (Evans, 1997). The eDNA results are in line with this observation that the two large rivers determine the limnological conditions in the lake (Beadle, 1981; Nakiyende et al., 2023). Rivers carry nutrients into lakes and strongly influence primary productivity at the river-lake transition zones (Cotte et al., 2023). This selects for species which can rapidly utilize nutrients for growth (Nhu et al., 2019). Localities with less nutrients often support specialized, highly efficient nutrient utiliser phytoplankton species which are competitive at low trophicity (Atkins et al., 2021).

eDNA recorded only a few clearly benthic taxa, represented by relatively few reads. This is expected as we sampled surface water. Consequently, most DNA fragments should originate from pelagic organisms. The macrobenthic fauna of Lake Albert is not well studied, but general records hint at a comparatively rich fauna (Guignot, 1957; Verbeke, 1957), representing groups such as Chironomidae, Ephemeroptera, and Ostracoda. Insects alone are estimated to number at least 200 species (Green, 2009). Green (2009) attributed the relative species richness of Lake Albert to the presence of sandpits along the Ugandan shores, which creates habitat heterogeneity. Chironomids were relatively abundant at Ntoroko in the eDNA data, but Ephemeroptera and ostracods were detected as well. eDNA supported the absence of other benthic taxa, such as decapod crustaceans which were so far not observed in the lake (Cumberlidge and Clark, 2018). A prominent group of benthic organisms are the molluscs, with the lake's fauna consisting of approximately 13 bivalve and 18 gastropod taxa (Brown, 1994; Mandahl Barth, 1954). eDNA recorded several mollusc ASVs such as

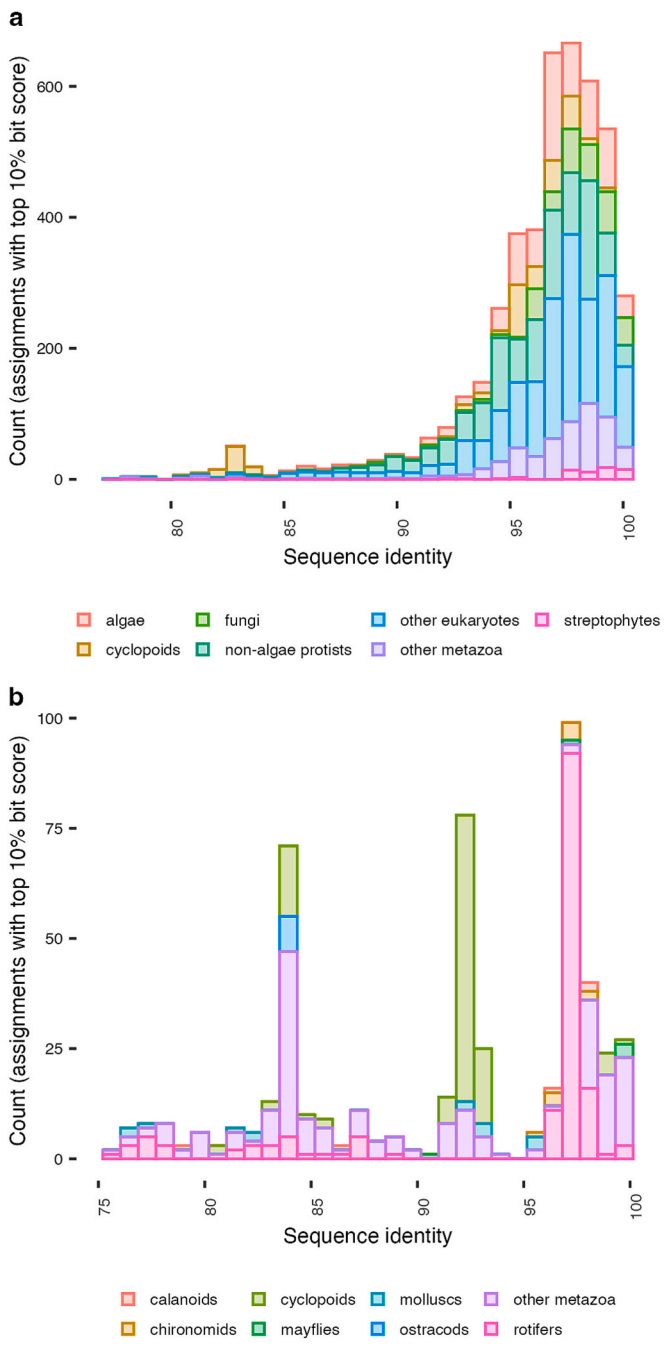


Fig. 4. Sequence identity of BLAST query and database sequences. a) Eukaryotes recorded with 18S. b) Metazoans recorded with COI.

sphaeriid bivalves (*Pisidium* sp.). Members of that genus are known to occur in Lake Albert (Mandahl Barth, 1954).

4.1.2. Differences in shore vs open water

Both markers showed statistically significant differences in community composition along a distance gradient from the shore toward the more central areas of the lake. The 18S dataset also showed a decrease in ASV richness from the shoreline toward the lake center. Read abundance of individual ASVs of both markers mostly decreased from the shore toward pelagic environments (Supp. Figs. 3–4) except one ASV assigned to diatoms. This is in agreement with expectations about differences between shoreline and central parts of large lakes, as nutrient concentrations generally decline from the shoreline to offshore areas. For example, in Lake Ontario, mean nutrient concentrations rapidly decline with increasing distance from the shore to background levels (Chomicki et al., 2022). Shoreline erosion can be an important source of nutrients, influencing lake productivity and trophic status (Hewlett et al., 2015). Water mixing can also affect nutrient availability, with more dynamic mixing of shallow shoreline waters promoting higher nutrient concentrations due to sediment resuspension (MacIntyre et al., 1999). Using traditional sampling, a steep decline in oxygen levels was observed with increasing depth, which has an impact on species diversity and richness (Dusabe et al., 2024). We collected our samples at a time of the year that is known for pronounced circulation of nutrients (Talling, 1965). Highly heterogeneous conditions at the shoreline, e.g. supported by macrophytes, also promote more taxonomically rich and compositionally distinct communities (Wang et al., 2018). Finally, reduction in oxygen in the pelagic zones, as observed in other African Great Lakes (Dusabe et al., 2024), might also play a role. The observed compositional gradient might also be a result of nutrient enrichment resulting from human activity, as all transects started at settlements. Many more locations will be necessary to disentangle the causes of the observed compositional gradients. A new sampling designed to disentangle these causes should contrast gradients starting at human settlements with gradients starting at shores without human impact. Vegetation conditions and the presence/absence of water inflows should be also considered when designing such a study in the future. While a designated sampling will be necessary to distinguish among the relative contribution of these effects to the observed patterns, eDNA records confirmed expectations about compositional differences between shoreline and pelagic communities.

4.2. Taxonomic identification of tropical eDNA sequences

The 18S sequences had an average identity of 96 % with sequences deposited in GenBank (Fig. 4). An average high identity with GenBank sequences is expected since 18S is a relatively conserved taxonomic marker, suitable for studying higher-level taxonomic units (Bininda-Emonds, 2021). However, many of the 18S ASVs (34 %) could not be assigned to taxonomic levels beyond eukaryotes. This might be caused by three processes. First, the query sequence might belong to an unknown eukaryotic group not present in the database. Second, the sequence in the database is misidentified, e.g., it is assigned to an organism but actually originates from a symbiont of this organism or contamination (Valkiūnas et al., 2008). Third, some of the assignments considered by the LCA algorithm are very low level, e.g. eukaryotes, which frequently happens when environmental sequences are deposited in NCBI. It is difficult to clearly distinguish among these three processes. Issues with reference sequence databases are common (summarized e.g. by (Chorlton, 2024). Issues with the identification quality of sequences deposited in NCBI certainly play a role (Valkiūnas et al., 2008). Nonetheless, taxonomic assignments in GenBank are fairly reliable (Leray et al., 2019), often comparable in quality to metazoan assignments made with BOLD (Baena-Bejarano et al., 2023). For the 18S marker gene we found a mismatch between the bit score (reflecting alignment length, sequence identity and database size) and sequence identity (Supp. Fig. 5). The vast majority of 18S ASV assignments were based on fairly low bit scores, regardless of the sequence identity, although all high bit score assignments also had high identity. This results from a short alignment length between query and database sequences, consistent with the level of conservation of the 18S: the ~150 bp V7 region amplified by us has a 20–25 bp highly variable core, flanked by conserved regions (Hadziavdic et al., 2014). Consequently, the most likely explanation for the large number of unidentified 18S eukaryotic ASVs is a general lack of eukaryotic sequences from African Great Lakes in GenBank.

Of the 1524 COI ASVs, 509 (33 %) could be assigned to metazoans. On average, these ASVs showed 91 % sequence identity with sequences deposited in GenBank and BOLD. Many other eukaryotes, including algae, fungi, and protists, were also amplified. This is not surprising, as

degenerate metazoan COI primers frequently amplify various groups beyond metazoans. Similarly to what we found here, metazoans might not even be the group with the highest read count (Collins et al., 2019). We analyzed only metazoans because the primers were optimized to specifically address primer biases for metazoans (Leray et al., 2013). We found a clear, almost linear relationship between identity and bit score (Supp. Fig. 5), which results from the highly conserved length of COI across the tree of life (Guo et al., 2022), as variable sites in COI are generally the second and third codon positions. Sequence identity of COI query and database sequences ranged between 75 and 100 %, with several distinct peaks (Fig. 4). Two prominent groups in the COI data are arthropods and rotifers. Arthropod query sequences had three distinct identity peaks: ~84 % (diverse metazoans), ~93 % (mostly cyclopoid copepods), and ~ 97–98 % (mostly rotifers). Rotifer COI queries had a distinct peak with most ASVs showing 97 % sequence identity with GenBank sequences. These ASVs were assigned to two cosmopolitan *Polyarthra* species (*P. dolichoptera* and *P. vulgaris*), both known for high levels of cryptic diversity even within the same water bodies (Liang et al., 2022; Obertegger et al., 2014). The two species accounted for >50 % of all rotifer counts in the COI dataset and >10 % of all metazoan counts. Both species were already reported from Lake Albert (Kiggundu et al., 2022; Lopez et al., 2023). Altogether, COI assignments results mirror those based on 18S: most sequenced COI fragments originate from species, genera, and families which do not have reference COI sequences available in GenBank. This is not surprising considering that many genera in well-researched regions such as North America also lack publicly available COI barcodes (Curry et al., 2018).

4.3. Management suggestions

Understanding of aquatic resources is crucial to inform policy formulation and decision-making. Currently, institutions such as NAFIRRI (National Fisheries Resources Research Institute), which hold the mandate for monitoring and conducting aquatic biodiversity research in Uganda, often omit important microbiota, such as zooplankton and phytoplankton, in their monitoring efforts due to limited resources. Most existing biodiversity data from Lake Albert are focused on fish, neglecting other critical organisms like zooplankton and algal communities, vital components of the food web and essential indicators of ecosystem health. Our study helps fill this gap by providing baseline data especially on the planktonic community composition of Lake Albert. Such baseline data is essential to support monitoring and management efforts, particularly in light of ongoing oil exploration activities.

Although our results show that the precise taxonomic assignment of eDNA sequences is problematic due to lack of reference sequence data, the results also highlight that aquatic eDNA is already suitable for taxonomy-free biomonitoring in the tropics. We recommend the urgent setup of an Africa-wide taxonomy-free eDNA-based monitoring scheme across freshwaters of the continent. Taxonomy-free biomonitoring directly links sequence data to environmental conditions and/or human impact, and allows large-scale DNA-based biomonitoring efforts (Cordier et al., 2019; Mächler et al., 2021). An eDNA-based, taxonomy-free biomonitoring scheme could efficiently track biodiversity dynamics by collecting aquatic eDNA samples at strategically placed locations at regular intervals. Such a scheme would be based on variation in α-, β- and γ-diversity of amplified sequence variants. The scheme would also allow the identification of bioindicator sequence variants, useful to authorities to quickly detect ecosystem disturbances. Further, a taxonomy-free eDNA monitoring scheme could be easily extended and refined to include economically important species and pathogens, highly relevant for societies across the continent. Systematic biomonitoring will be particularly important for managing local and large-scale threats, such as pollution, resource overuse, species invasions, climate change, allowing for interventions to ensure resilience.

5. Conclusions and outlooks

In summary, our study confirmed that aquatic eDNA is suitable to record biodiversity patterns in large, warm tropical freshwaters. Environmental DNA reflected both previous and expected findings about the structure and composition of planktonic communities in Lake Albert, in support of the first and second study hypotheses. Environmental DNA revealed significant differences in community composition in distinct localities, confirming previous, morphology-based findings about processes that shape planktonic communities. The eDNA data, consistent with limited morphology-based observations, indicated that cyclopoid copepods dominate pelagic communities, controlling planktonic producers. The data also confirmed previous findings about the importance of riverine nutrient input as the primary factor influencing planktonic microorganisms in the lake, confirming earlier observations. Environmental DNA was sufficiently sensitive to reveal gradients in plankton community composition. These gradients were expected both from historic observations, and on the basis of limnological differences between shoreline and pelagic regions.

However, the current dataset only included pelagic zones, neglecting the benthic regions and the entire water column. This limitation, along with the restricted spatial and temporal scope of the data, highlights the need for more comprehensive spatial experimental designs. These should include replicated sampling at both riverine and non-riverine sites, and transects starting from both settlements and natural areas. When designing these studies, we recommend considering further limnological factors specific to tropical lakes. Several of these factors, such as wind mixing together with a weak thermal stratification of the water column (Lewis, 1987) may strongly influence community composition of the plankton, as oxygen and nutrient conditions might be homogenized vertically and horizontally. This may result in more homogeneous planktonic communities, requiring longer gradients or more samples to disentangle compositional patterns. Furthermore, seasonal aspects, such as dry and wet seasons, and temporal variations in planktonic composition and abundance should be considered.

The results of the present study clearly confirmed expected strong patterns in the distribution of the most common taxa, even with a relatively limited sample set from three transects at strategically selected locations. However, investigating rarer taxa, particularly those with restricted distributions within the lake, would benefit from sampling at a much higher spatial resolution. Future studies should consider the shoreline of Lake Albert in the Democratic Republic of Congo. Ideally, such efforts would involve international teams from both countries. Sampling along the entire shoreline will undoubtedly increase the number of identified taxa and significantly contribute to our understanding of high-resolution patterns in planktonic community composition. The results also confirmed the third hypothesis: although eDNA data reveals strong biodiversity patterns, a major challenge to its application in tropical regions like Africa is the urgent need for generating comprehensive reference sequence data. The lack of reference sequences significantly hampers the ability to accurately interpret eDNA results from the most biodiverse regions of Earth, as many taxa remain unidentified or misidentified due to insufficient database coverage. Without reference sequences, eDNA can record ecological patterns but struggles with accurate interpretation. This is because eDNA results cannot be effectively linked to existing knowledge about taxa, their ecology, habitat preferences, and traits. This gap underscores the necessity for coordinated efforts to expand genetic databases with sequences from diverse taxa present in tropical freshwater ecosystems. We recommend prioritizing potentially high biomass species when generating reference sequence data, as these species likely have a large impact on ecosystem functioning.

Despite these challenges, we advocate for eDNA as a highly suitable monitoring tool in tropical regions, even in the current absence of reliable marker gene data. Our approach underscores the importance of targeting entire ecosystems, including both producers and consumers,

rather than focusing on specific groups. It is crucial to exploit the capabilities of molecular tools for comprehensive surveys in highly diverse tropical ecosystems. Only by doing this can we achieve a holistic understanding of ecosystems and the complex interaction networks among microorganisms, plants, animals, fungi, and protists, and perform efficient interventions to ensure their resilience.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.177308.

CRediT authorship contribution statement

Miklós Bálint: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Julius Tumusiime:** Writing – review & editing, Investigation. **Justine Nakintu:** Writing – review & editing, Investigation. **Damian Baranski:** Writing – review & editing, Methodology. **Leonie Schardt:** Writing – review & editing, Methodology. **Juliane Romahn:** Writing – review & editing, Methodology, Data curation. **Marie-Claire Dusabe:** Writing – review & editing, Investigation. **Casim Umba Tolo:** Writing – review & editing, Investigation. **Grace Rugunda Kagoro:** Writing – review & editing, Investigation. **Francis Ssenkuba:** Writing – review & editing, Investigation. **Annett Junginger:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Conceptualization. **Christian Albrecht:** Writing – review & editing, Writing – original draft, Project administration, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT v4 in order to improve the readability and language of the manuscript and to optimize data analysis scripts. The authors also used Perplexity for literature research. After using these tools, the authors reviewed and edited the content as needed. They take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are thankful to Mbarara University of Science and Technology (MUST) for providing staff time to participate in the KEA-EXPLORE. We are grateful to the local people of the KEA-EXPLORE pilot project areas in the three districts of Uganda (Ntoroko, Hoima, and Buliisa) for helping us during the fieldwork, especially the local leaders and the owners of the canoes used during sample collection on the Lake Albert. We appreciate the reviews and guidance made by the Institutional Review Board of Mbarara University of Science and Technology which enabled us to get research approval and material transfer clearance from the Uganda National Council for Science and Technology. We thank Arthur Boom and Hugo Gante (Royal Museum for Central Africa, Belgium) for discussions on the bioinformatics and analysis of metabarcoding data. This work was supported by the "Senckenberg Green Heart of Africa" seed funds, and the LOEWE Program of the Hessian Ministry of Higher Education, Research and the Arts (LOEWE/1/10/519/03/03.001(0014)/52).

Data availability

High throughput sequence data will be accessible in NCBI or ENA upon acceptance. Analyses and results can be reproduced with R scripts and data deposited on FigShare (doi: https://doi.org/10.6084/m9.figshare.26506030).

References

Ali, M.M., Abd Ellah, R.G., 2023. Chapter 1 - history and formation of African Lakes. In: El-Sheekh, M., Elsaied, H.E. (Eds.), Lakes of Africa. Elsevier, pp. 1–31.
Atkins, K.S., Hackley, S.H., Allen, B.C., Watanabe, S., Reuter, J.E., Schladow, S.G., 2021. Variability in periphyton community and biomass over 37 years in Lake Tahoe (CA-NV). Hydrobiologia 848, 1755–1772.
Baena-Bejarano, N., Reina, C., Martínez-Revelo, D.E., Medina, C.A., Tovar, E., Uribe-Soto, S., Neita-Moreno, J.C., Gonzalez, M.A., 2023. Taxonomic identification accuracy from BOLD and GenBank databases using over a thousand insect DNA barcodes from Colombia. PloS One 18, e0277379.
Bairoliya, S., Koh Zhi Xiang, J., Cao, B., 2022. Extracellular DNA in environmental samples: occurrence, extraction, quantification, and impact on microbial biodiversity assessment. Appl. Environ. Microbiol. 88, e01845-21. https://doi.org/10.1128/aem.01845-21.
Bálint, M., Bahram, M., Eren, A.M., Faust, K., Fuhrman, J.A., Lindahl, B., O'Hara, R.B., Öpik, M., Sogin, M.L., Unterseher, M., Tedersoo, L., 2016. Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. FEMS Microbiol. Rev. 40, 686–700.
Bálint, M., Márton, O., Schatz, M., Düring, R.-A., Grossart, H.-P., 2018. Proper experimental design requires randomization/balancing of molecular ecology experiments. Ecol. Evol. 8, 1786–1793.
Beadle, L.C., 1981. The Inland Waters of Tropical Africa: An Introduction to Tropical Limnology. Longman.
Beauchamp, R.S.A., 1956. The electrical conductivity of the head-waters of the white Nile. Nature 178, 616–619.
Berke, M.A., Johnson, T.C., Werne, J.P., Livingstone, D.A., Grice, K., Schouten, S., Sinninghe Damsté, J.S., 2014. Characterization of the last deglacial transition in tropical East Africa: insights from Lake Albert. Palaeogeogr. Palaeoclimatol. Palaeoecol. 409, 1–8.
Beuning, K.R.M., Talbot, M.R., Kelts, K., 1997. A revised 30,000-year paleoclimatic and paleohydrologic history of Lake Albert, East Africa. Palaeogeogr. Palaeoclimatol. Palaeoecol. 136, 259–279.
Bininda-Emonds, O.R.P., 2021. 18S rRNA variability maps reveal three highly divergent, conserved motifs within Rotifera. BMC Ecol Evol 21, 118.
Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. obitools: a unix-inspired software package for DNA metabarcoding. Mol. Ecol. Resour. 16, 176–182.
Brown, D.S., 1994. Freshwater Snails of Africa and Their Medical Importance. CRC Press.
Chomicki, K.M., Taylor, W.D., Brown, C.J.M., Dove, A., Bowen, G.S., Mohamed, M.N., 2022. Seasonal variation in the influence of environmental drivers on nearshore water quality along an urban northern Lake Ontario shoreline. J. Great Lakes Res. 48, 914–926.
Chorlton, S.D., 2024. Ten common issues with reference sequence databases and how to mitigate them. Front Bioinform 4, 1278228.
Collins, R.A., Bakker, J., Wangensteen, O.S., Soto, A.Z., Corrigan, L., Sims, D.W., Genner, M.J., Mariani, S., 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods Ecol. Evol. 10, 1985–2001.
Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., Pawlowski, J., 2019. Embracing environmental genomics and machine learning for routine biomonitoring. Trends Microbiol. 27, 387–397.
Cotte, G., Soulignac, F., dos Santos Correia, F., Fallet, M., Ibelings, B.W., Barry, D.A., Vennemann, T.W., 2023. Controlling factors of phytoplankton distribution in the river–lake transition zone of a large lake. Aquat. Sci. 85, 37.
Cumberlidge, N., Clark, P.F., 2018. Albertine Rift Valley endemics: three new species of freshwater crabs (Brachyura: Potamoidea: Potamonautidae) from Uganda, the Democratic Republic of the Congo, and Rwanda. J. Nat. Hist. 52, 1637–1656.
Curry, C.J., Gibson, J.F., Shokralla, S., Hajibabaei, M., Baird, D.J., 2018. Identifying North American freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? Freshw. Sci. 37, 178–189.
De Crop, W., Verschuren, D., 2021. Mixing regimes in the equatorial crater lakes of western Uganda. Limnologica 90, 125891. https://doi.org/10.1016/j.limno.2021.125891.
Doble, C.J., Hipperson, H., Salzburger, W., Horsburgh, G.J., Mwita, C., Murrell, D.J., Day, J.J., 2020. Testing the performance of environmental DNA metabarcoding for surveying highly diverse tropical fish communities: a case study from Lake Tanganyika. Environ. DNA 2, 24–41.
Drake, L.E., Cuff, J.P., Young, R.E., Marchbank, A., Chadwick, E.A., Symondson, W.O.C., 2022. An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data. Methods Ecol. Evol. 13, 694–710.
Dusabe, M.C., Kalinda, C., Clewing, C., Hyangya, B.L., Van Bocxlaer, B., Albrecht, C., 2024. Environmental perturbations and anthropogenic disturbances determine mollusc biodiversity of Africa's explosive Lake Kivu. J. Great Lakes Res. 50, 102339.
Ericson, K., 2014. A Crude Awakening: The Relationship between Petroleum Exploration and Environmental Conservation in Western Uganda. Collection, Independent Study Project (ISP).
Evans, J.H., 1997. Spatial and seasonal distribution of phytoplankton in an African Rift Valley lake (L. Albert, Uganda, Zaire). Hydrobiologia 354, 1–16.
Froese, R., Pauly, D., 2024. FishBase. World Wide Web electronic publication [WWW Document]. https://fishbase.de/ (accessed 10.11.24).

Geller, J., Meyer, C., Parker, M., Hawk, H., 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Mol. Ecol. Resour. 13, 851–861.
Green, J., 1967. The distribution and variation of *Daphnia lumholtzi* (Crustacea: Cladocera) in relation to fish predation in Lake Albert. East Africa. J. Zool. 1987 (151), 181–197.
Green, J., 1971. Associations of Cladocera in the zooplankton of the lake sources of the White Nile. J. Zool. 1987 (165), 373–414.
Green, J., 2009. Nilotic Lakes of the Western rift. In: Dumont, H.J. (Ed.), The Nile: Origin, Environments, Limnology and Human Use. Springer, Netherlands, Dordrecht, pp. 263–286.
Guardiola, M., Uriz, M.J., Taberlet, P., Coissac, E., Wangensteen, O.S., Turon, X., 2015. Deep-Sea, deep-sequencing: Metabarcoding extracellular DNA from sediments of marine canyons. PloS One 10, e0139633.
Guignot, F., 1957. Haliplidae et Dytiscidae. Exploration hydrobiologique des lacs Kivu, Edouard et Albert 3, 131–169.
Guo, M., Yuan, C., Tao, L., Cai, Y., Zhang, W., 2022. Life barcoded by DNA barcodes. Conserv. Genet. Resour. 14, 351–365.
Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., Troedsson, C., 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PloS One 9, e87624.
Hewlett, C., North, R.L., Johansson, J., Vandergucht, D.M., Hudson, J.J., 2015. Contribution of shoreline erosion to nutrient loading of the Lake Diefenbaker reservoir, Saskatchewan. Canada. J. Great Lakes Res. 41, 110–117.
Hui, F.K.C., 2016. Boral - Bayesian ordination and regression analysis of multivariate abundance data in r. Methods Ecol. Evol. 7, 744–750.
Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., 2016. MEGAN Community edition - interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput. Biol. 12, e1004957.
IPBES, 2019. Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat, Bonn, Germany.
Kiggundu, V., Egessa, R., Mwebaza-Ndawula, L., 2022. Zooplankton Occurrences and Abundance in the Lake Albert System, Uganda. https://doi.org/10.15468/WQ98M3.
Krehenwinkel, H., Wolf, M., Lim, J.Y., Rominger, A.J., Simison, W.B., Gillespie, R.G., 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Sci. Rep. 7, 17668.
Lamb, P.D., Fonseca, V.G., Maxwell, D.L., Nnanatu, C.C., 2022. Systematic review and meta-analysis: Water type and temperature affect environmental DNA decay. Mol. Ecol. Resour. 22, 2494–2505. https://doi.org/10.1111/1755-0998.13627.
Lehman, J.T., Litt, A.H., Mugidde, R., Lehman, D.A., 1998. Nutrients and plankton biomass in the rift Lake sources of the White Nile: Lakes Albert and Edward. In: Lehman, J.T. (Ed.), Environmental Change and Response in East African Lakes. Springer, Netherlands, Dordrecht, pp. 157–172.
Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., Machida, R.J., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front. Zool. 10, 34.
Leray, M., Knowlton, N., Ho, S.-L., Nguyen, B.N., Machida, R.J., 2019. GenBank is a reliable resource for 21st century biodiversity research. Proc. Natl. Acad. Sci. U. S. A. 116, 22651–22656.
Lewis, W.M., 1987. Tropical Limnology. Annu. Rev. Ecol. Syst. 18, 159–184.
Liang, D., McManus, G.B., Wang, Q., Sun, X., Liu, Z., Lin, S., Yang, Y., 2022. Genetic differentiation and phylogeography of rotifer Polyarthra dolichoptera and P. vulgaris populations between Southeastern China and eastern North America: high intercontinental differences. Ecol. Evol. 12, e8912.
Lindén, A., Mäntyniemi, S., 2011. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology 92, 1414–1421.
Lopes-Lima, M., Prié, V., Camará, M., Ceríaco, L.M.P., Fernandes, V., Ferreira, S., Gonçalves, D.V., Lecoq, M., Martins, F.M.S., Parrinha, D., de Barros, A.R., Valentini, A., Veríssimo, J., Palma, L., Beja, P., 2024. Rapid eDNA survey reveals a unique biodiversity hotspot: the Corubal River, West Africa. BioScience 74, 405–412. https://doi.org/10.1093/biosci/biae036.
Lopez, Z.F., Cancellario, T., Fontaneto, D., Kamburska, L., Karimullah, K., Wallace, R.L., Walsh, E.J., Smolak, R., 2023. African Rotifer Records. https://doi.org/10.15468/VE5YGW.
Mächler, E., Walser, J.-C., Altermatt, F., 2021. Decision-making and best practices for taxonomy-free environmental DNA metabarcoding in biomonitoring using Hill numbers. Mol. Ecol. 30, 3326–3339.
MacIntyre, S., Flynn, K.M., Jellison, R., Romero, J.R., 1999. Boundary mixing and nutrient fluxes in Mono Lake. California. Limnol. Oceanogr. 44, 512–529.
Mandahl Barth, G., 1954. The freshwater mollusks of Uganda and adjacent territories. In: Annales du Musée Royal du Congo Belge, Sciences Zoologiques, 32, pp. 1–206.
Meglécz, E., 2023. COInr and mkCOInr: building and customizing a nonredundant barcoding reference database from BOLD and NCBI using a semi-automated pipeline. Mol. Ecol. Resour. 23, 933–945.
Morana, C., Borges, A.V., Deirmendjian, L., Okello, W., Sarmento, H., Descy, J.-P., Kimirei, I.A., Bouillon, S., 2023. Prevalence of autotrophy in non-humic African Lakes. Ecosystems 26, 627–642.
Nakiyende, H., Basooma, A., Ikwaput Nyeko, J., Okello, W., Rugadya, R., Albrecht, C., Lawrence, T., Van Steenberge, M., Smith, S., Muderhwa, N., Matunguru, J., Mulongaibalu, M., Ajode, M.Z., 2023. Limitations for informed decision making and better management of the transboundary Lake Albert fisheries resources. J. Great Lakes Res. 49, 102165.
Nhu, Y.D.T., Hoang, N.T., Lieu, P.K., Harada, H., Brion, N., Van Hieu, D., Van Hop, N., Olde Venterink, H., 2019. Effects of nutrient supply and nutrient ratio on diversity-productivity relationships of phytoplankton in the Cau Hai lagoon. Vietnam. Ecol. Evol. 9, 5950–5962.
Obertegger, U., Flaim, G., Fontaneto, D., 2014. Cryptic diversity within the rotifer *Polyarthra dolichoptera* along an altitudinal gradient. Freshw. Biol. 59, 2413–2427.
Odada, E.O., Olago, D.O., 2006. The East African Great Lakes: Limnology. Springer Science & Business Media, Palaeolimnology and Biodiversity.
Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., Abrego, N., 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol. Lett. 20, 561–576.
R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rowel, C., Fred, B., Betson, M., Sousa-Figueiredo, J.C., Kabatereine, N.B., Stothard, J.R., 2015. Environmental epidemiology of intestinal schistosomiasis in Uganda: population dynamics of biomphalaria (gastropoda: planorbidae) in Lake Albert and Lake Victoria with observations on natural infections with digenetic trematodes. Biomed. Res. Int. 2015, 717261.
Salzburger, W., Van Bocxlaer, B., Cohen, A.S., 2014. Ecology and evolution of the African Great Lakes and their faunas. Annu. Rev. Ecol. Evol. Syst. 45, 519–545.
Schenekar, T., 2023. The current state of eDNA research in freshwater ecosystems: are we shifting from the developmental phase to standard application in biomonitoring? Hydrobiologia 850, 1263–1282.
Sengupta, M.E., Lynggaard, C., Mukaratirwa, S., Vennervald, B.J., Stensgaard, A.S., 2022. Environmental DNA in human and veterinary parasitology - current applications and future prospects for monitoring and control. Food Waterborne Parasitol 29, e00183.
Simpson, G.L., 2018. Modelling palaeoecological time series using generalised additive models. Front. Ecol. Evol. 6. https://doi.org/10.3389/fevo.2018.00149.
Taberlet, P., Bonin, A., Zinger, L., Coissac, E., 2018. Environmental DNA: For Biodiversity Research and Monitoring, Illustrated Edition. ed. Oxford Univ Pr.
Takahashi, M., Saccò, M., Kestel, J.H., Nester, G., Campbell, M.A., van der Heyde, M., Heydenrych, M.J., Juszkiewicz, D.J., Nevill, P., Dawkins, K.L., Bessey, C., Fernandes, K., Miller, H., Power, M., Mousavi-Derazmahalleh, M., Newton, J.P., White, N.E., Richards, Z.T., Allentoft, M.E., 2023. Aquatic environmental DNA: a review of the macro-organismal biomonitoring revolution. Sci. Total Environ. 873, 162322.
Talling, J.F., 1963. Origin of stratification in an African rift lake. Limnol. Oceanogr. 8, 68–78.
Talling, J., 1965. The photosynthetic activity of phytoplankton in East African lakes. Int. Rev. Hydrobiol. 50, 1–32.
Talling, J.F., 1969. The incidence of vertical mixing, and some biological and chemical consequences, in tropical African lakes. SIL Proc. 17, 998–1012. https://doi.org/10.1080/03680770.1968.11895946.
Tumusiime, J., Bhraunxs, N.V., Kagoro-Rugunda, G., Namirembe, D., Albrecht, C., Twongyirwe, R., Tolo, C.U., Jacobs, L., Huyse, T., 2024. Citizens can help to map putative transmission sites for snail-borne diseases. PLoS Negl. Trop. Dis. 18, e0012062.
Valkiūnas, G., Atkinson, C.T., Bensch, S., Sehgal, R.N.M., Ricklefs, R.E., 2008. Parasite misidentifications in GenBank: how to minimize their number? Trends Parasitol. 24, 247–248.
Verbeke, J., 1957. Recherches écologiques sur la faune des grands lacs de l'est du Congo belge. Bulletin de l'Institut royal des Sciences naturelles de Belgique: Résultats scientifiques de l'exploration hydrobiologique (1952-1954) des lacs Kivu, Edouard et Albert. - Bulletin van het Koninklijk Belgisch Instituut voor Natuurwetenschappen: Wetenschappelijke resultaten van exploratie hydrobiologische (1952-1954) meren van Kivu. Edward en Albert. III, 3–177.
Verheyen, E., Abila, R., Akoll, P., Albertson, C., Antunes, D., Banda, T., Bills, R., Bulirani, A., Manda, A.C., Cohen, A.S., Cunha-Saraiva, F., Derycke, S., Donohue, I., Du, M., Dudu, A.M., Egger, B., Fritzsche, K., Frommen, J.G., Gante, H.F., Genner, M. J., Härer, A., Hata, H., Irvine, K., Mwapu, P.I., de Bisthoven, L.J., Jungwirth, A., Kaleme, P., Katongo, C., Kéver, L., Koblmüller, S., Konings, A., Lamboj, A., Lemmel-Schaedelin, F., Schiaffino, G.M., Martens, K., Mulungula, P.M., Meyer, A., More, H. L., Musilova, Z., Bukinga, F.M., Muzumani, R., Ntakimazi, G., Okello, W., Phiri, H., Pialek, L., Plisnier, P.D., Raeymaekers, J.A.M., Rajkov, J., Rican, O., Roberts, R., Salzburger, W., Schoen, I., Sefc, K.M., Singh, P., Skelton, P., Snoeks, J., Schneider, K., Sturmbauer, C., Svardal, H., Svensson, O., Dowdall, J.T., Turner, G.F., Tyers, A., van Rijssel, J.C., Van Steenberge, M., Vanhove, M.P.M., Weber, A.-T., Weyl, O., Ziegelbecker, A., Zimmermann, H., 2016. Oil extraction imperils Africa's Great Lakes. Science 354, 561–562.
von der Heyden, S., 2023. Environmental DNA surveys of African biodiversity: state of knowledge, challenges, and opportunities. Environ. DNA 5, 12–17.
von der Heyden, S., Neef, G., Grevesse, T., Cwecwe, Y., Sado, T., Miya, M., Mosie, I., Creer, S., Skelton, P., von Brandis, R., 2023. Environmental DNA biomonitoring in biodiversity hotspots: a case study of fishes of the Okavango Delta. Environ. DNA 5, 1720–1731.
Wandera, S.B., Balirwa, J.S., 2010. Fish species diversity and relative abundance in Lake Albert—Uganda. Aquat. Ecosyst. Health Manag. 13, 284–293.
Wang, S.C., Liu, X., Liu, Y., Wang, H., 2018. Contrasting patterns of macroinvertebrates inshore vs. offshore in a plateau eutrophic lake: implications for lake management. Limnologica 70, 10–19.

Wang, Y., Naumann, U., Wright, S.T., Warton, D.I., 2012. mvabund– an R package for model-based analysis of multivariate abundance data. Methods Ecol. Evol. 3, 471–474.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., Hui, F.K.C., 2015. So many variables: joint modeling in community ecology. Trends Ecol. Evol. 30, 766–779.

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R, Second edition. CRC Press.

View publication stats